

Kathleen L. Mosier, Dietrich Manzey

Humans and Automated Decision Aids: A Match Made in Heaven?

Subtitle

Chapter in book | Accepted manuscript (Postprint)

This version is available at <https://doi.org/10.14279/depositonce-10992>



Mosier, K. L., & Manzey, D. (2019). Humans and Automated Decision Aids: A Match Made in Heaven? Human Performance in Automated and Autonomous Systems, 19–42.
<https://doi.org/10.1201/9780429458330-2>

This is an Accepted Manuscript of a book chapter published by CRC Press in Human Performance in Automated and Autonomous Systems on 19 September 2019, available online:
<https://doi.org/10.1201/9780429458330-2>

Terms of Use

Copyright applies. A non-exclusive, non-transferable and limited right to use is granted. This document is intended solely for personal, non-commercial use.

WISSEN IM ZENTRUM
UNIVERSITÄTSBIBLIOTHEK

Technische
Universität
Berlin

Humans and Automated Decision Aids: A Match Made in Heaven?

Kathleen L. Mosier
San Francisco State University

Dietrich Manzey
Technical University Berlin

In the tiny villages and towns around Winchester, UK, motorists have long had problems with GPS navigation. They follow the route their devices suggest and repeatedly wind up stuck - the roads are simply too narrow for some wider vehicles. The problem got so bad in the town of Exton that a bright yellow [sign was erected](https://www.ranker.com/list/9-car-accidents-caused-by-google-maps-and-gps/robert-wabash) to warn drivers NOT to follow their GPS and to, instead, rely on common sense. These are the times we live in, folks. (<https://www.ranker.com/list/9-car-accidents-caused-by-google-maps-and-gps/robert-wabash>)

Automated Decision Support: What and Why?

For the last four decades, computer-based automated or even autonomous systems have changed our life considerably. A specific sort of automation is so called Automated Decision Support Systems (DSSs), which are ubiquitous in today's high-technology world. Designed as a support tool for humans, they provide at discrete points in time, either automatically or on demand, certain information about the state of the world that can improve informed decision-making. Beyond that, they also may provide recommendations on actions and/or predictions of outcomes in the near future in order to help the human to arrive at proper decisions about actions. Their main goal is to improve the reliability and quality of human decision-making at work or in everyday life.

Two basic aspects of human decision-making have been referred to as *front-end* and *back-end* processes (Mosier & Fischer, 2010). *Front-end* processes include all cognitive judgement processes needed for a proper situation assessment and situation awareness (SA) (Endsley, 1995, 2000), for example information search, information analysis, risk assessment or identification of constraints for possible actions. These usually provide the basis for the *back-end* processes which include the actual decision-making, that is, the selection and planning of proper actions. Corresponding to this basic distinction, DSSs can be categorized, albeit relatively coarsely, as to whether they support the *front-end* or *back-end* of decision-making. Typical examples of front-end DSSs are alarm systems, such as alarm

lights that illuminate on cockpit or other control displays to alert the operator about a problem, or the Collision Prevention Assist in an automobile, a sensor that monitors a vehicle's distance from vehicles in front and beeps if the driver accidentally gets too close to the car ahead. More complex *front-end* assistance systems include information systems, such as those that provide traffic information on radar displays or moving maps or provide surgeons with information about the location of specific landmarks to help them navigate through a patient's anatomy. In contrast, *back-end* DSSs include all systems that provide automated support for decision choice and action selection, such as the GPS systems in cars which analyze a given situation and then suggest the quickest route to a location or instruct the individual exactly on what to do. Even more complex systems may dynamically incorporate both *front-end* and *back-end* processes. Examples of this latter type are sophisticated DSSs in aircraft cockpit like the Ground Proximity Warning System (GPWS) or the Traffic Collision Avoidance System (TCAS) which start as pure *front-end* systems alerting pilots of a possibly critical situation (e.g. "terrain, terrain" or "traffic, traffic") but switch to *back-end* directives instructing the pilot exactly what to do if the situation becomes more critical (e.g. "pull up" or "climb/descend"). Note that our distinction of *front-end* and *back-end* DSSs also fits to models describing different types and levels or degrees of automation (e.g. Endsley & Kaber, 1999; Onnasch, Wickens, Li & Manzey, 2014; Parasuraman, Sheridan & Wickens, 2000). For example, Parasuraman et al. (2000) distinguish different types and levels of automation with types of automation referring to the stage of information processing that is supported. Automation support of the first two stages – information acquisition and analysis – would correspond to *front-end* support, while automation support of later stages of information processing, i.e. decision and response selection, would correspond to *back-end* support. Because back-end DSS often implicitly also include automated situation assessment they can be considered to represent a higher degree of automation than front-end support, that is they support the human user in both situation assessment and selection of actions (Onnasch et al., 2014).

The intent of all of these DSSs is to compensate for the inadequacies of the human decision maker and to make decisions more rational, compared to the often observed issues of biases and heuristics in

human decision making (Gigerenzer & Todd, 1999; Kahnemann, 2011). If this would work, providing a DSS as support tool for a human decision maker would result in a well-suited pairing with an almost perfect outcome - like a match made in heaven! However, before treating DSSs in such romantic way, some caveats are necessary. Evidence over the past 22 years since this volume was first published suggests that the relationship between human decision makers and DSSs is not as ideal as expected. Although much of the time DSSs successfully enhance human performance, reliance on imperfect automation has the unintended consequence of worsening performance when automated information or advice is not correct. In fact, human tendencies toward the use of heuristics, or short-cuts, in judgment and decision making, which DSSs should help to avoid, can be triggered by DSSs, eroding the quality of the human-automation match and causing it to be more earth-like with weaknesses as well as strengths than a heavenly ideal. In particular, the presence of DSSs may foster *automation bias*, “the tendency to use automated cues as a heuristic replacement for vigilant information seeking and processing...” (Mosier & Skitka, 1996, p. 205).

This chapter particularly focusses on the performance consequences of DSSs with a particular emphasis on the issue of automation bias as demonstrated in research across many domains. We examine the factors that impact automation bias including operator characteristics such as trust in the system, level of perceived accountability, system characteristics such as reliability, transparency and understandability, the level and degree of automation involved and the characteristics of the task-context including time-pressure, consequences of errors, workload or working in redundant work teams. We also discuss the challenge of balancing the pros and cons of using automation as a heuristic in decision making as well as current trends and future advances in automated system.

Automation Bias: *Still around after all these years.*

Designers responded to "pilot error" and the increasing cockpit workload by attempting to remove the error at its source, that is, to replace human functioning with device functioning; in their view, to automate human error out of the system. But there were two flaws in this reasoning. (1) The devices themselves had to be operated and monitored by the very humans whose caprice they were designed to avoid; human error was not eliminated, but relocated. (2)

The devices themselves had the potential for generating errors that could result in accidents.
(Wiener, 1985, p. 78)

Automation bias was first identified in cockpit crews and was characterized in the tradition of classical decision heuristics and biases (Kahneman, Slovic, & Tversky, 1982) as a decision short-cut. It manifests in two types of performance consequences: *omission errors* result when decision makers do not take appropriate action because they are not informed about a situation, problem, or imminent system failure by automated decision aids; *commission errors* occur when decision makers take inappropriate action because they over-attend to information or directions from an automated decision aid (Mosier & Skitka, 1996; Mosier, Skitka, Heers, & Burdick, 1998). The distinction between two manifestations of automation bias is related to the distinction of reliance and compliance in human-automation interaction proposed by Meyer (2004). Whereas reliance describes how much an individual trusts the alert function of a DSS in case of critical events and, thus, reduces his or her own monitoring of the environment, compliance describes the behavioral tendency to trust and follow an DSS's recommendation. Given this, automation bias in terms of omission errors can be understood as a consequence of overreliance on a DSS, whereas commission errors reflect a sort of overcompliance with a DSS.

Several possible mechanisms underlie these errors. First, automated aids present powerful, salient decisional cues and are widely believed to be accurate. The 'cognitive miser' perspective suggests that people tend to take the road of least cognitive effort and may diffuse responsibility for decision tasks to automated aids, resulting in incomplete cross-checking of available information. This is related to automation-induced complacency, which has been documented as a factor in automation-related incidents and accidents (see Parasuraman & Manzey, 2010, for a review). Second, operators may inattentively process information that contradicts automated directives, analogous to a looking but not seeing' or *inattentional blindness* effect (Simons & Chabris, 1999), or may 'see' information that confirms the automation even though it is not actually present, a phenomenon labelled 'phantom memory' (Mosier et al., 1998). A third possibility is that people see but actively discount information from less automated sources because they believe the automated decision support system is more likely correct (Skitka et al.,

2000). It is likely that the mechanism varies along with the task context, for example, whether automated information is made more salient than other information or whether time pressure is present.

Additionally, the mechanism may depend on whether the DSS targets the front or back end of decision making.

Since the early research in the 1990s, automation bias has been identified not only in aviation but also in several other domains in which technology and automated decision support are commonplace.

Evidence for Automation Bias from Different Domains.

Aviation. A series of studies of human-automation interaction in aviation (and aviation-like laboratory tasks) by Mosier and colleagues provided the cradle of *automation bias* research (Mosier, Palmer & Degani, 1992; Mosier, Skitka, Heers & Burdick, 1998; Skitka, Mosier & Burdick, 1999). In the first study of this series, Mosier et al. (1992) studied the effects an automated electronic checklist on decision-making in an ambiguous simulated engine fire situation, that is, there were conflicting cues concerning which engine was on fire. The electronic checklist for ENGINE FIRE prompted the pilots to shut down the wrong engine (i.e. the one not affected by fire). Although available information from other cockpit instruments contradicted the electronic device, a total of 75% of pilots followed the wrong recommendation of the decision-aid. This was in stark contrast to a control group of pilots who used a traditional checklist for the same task and committed this error to a considerably lesser extent (25%).

In another study a sample of experienced commercial pilots performed simulated flights in a low-fidelity flight trainer equipped with various advanced cockpit automation systems (Mosier et al., 1998). During the flight they were exposed to automation failures creating situations for omission and commission errors, including an altitude clearance that did not get loaded correctly, a heading change that was not executed properly by the flight system, and a false engine-fire alert on the electronic EICAS (Engine Indicating and Crew Alerting System) that was not supported by any other instrument readings. While the first two automation errors provided opportunities for an error of omission if not detected by the pilots, the latter one provided an opportunity to commit a commission error. Basically, all of these

automation failures were easily detectable by use of relevant information available from other cockpit instruments. Yet, the frequency of omission errors was 55% - and 100% of the pilots decided to shut down the engine in response to the erroneous EICAS alert. Interestingly, 67% of the pilots reported that they had seen at least one other indication supporting the EICAS fire alert which in fact was not there.

Further evidence for automation bias was also found in a group of non-pilots performing a low level flight simulation task with and without automation support (Skitka et al., 1999). During the task several automation failures occurred - six failures where the automation failed to prompt the participants of a critical event, and another six where the aid gave a wrong recommendation. Again the omission error rate in response to the first type of automation failure was 41%. In contrast only 3% of the critical events were missed by a control group performing the task without automation support. A similar picture also emerged for commission errors. A total of 65% of participants followed the wrong advice of the decision aid, although other available information directly contradicted the advice and the participants were informed that the other information would be perfectly valid. The common characteristic of all these findings was that the presence of an automated support for decisions made pilots less likely to seek and assess information from other sources. Since this early work, many other examples of automation bias in interaction with decision aids in advanced cockpits and also air traffic control have been reported (e.g. Metzger & Parasuraman, 2005; Rovira & Parasuraman, 2010; Sarter & Schroeder, 2001), also as contributing factors to fatal accidents (e.g. Dutch Safety Board, 2010). Many of these cases suggest that automation bias might not necessarily result in manifest errors of omission (or commission) but rather a delayed detection and response to contradictory information leaving insufficient time to respond to a critical situation (de Boer, Hems & Hurts, 2014).

Health Care. Automation bias in interactions with automated decision-aids has also become a concern in the health care domain (Goddard, Roudsari & Wyatt, 2012, 2014; Lyell et al., 2017). One set of studies addressed the performance consequences of so called clinical decision-support systems (CDSSs) on the quality of clinical judgement and decision-making (Goddard et al., 2014; Lyell et al., 2017; McKibbin & Fridsma, 2006; Tsai, Fridsma & Gatti, 2003; Westbrook, Coiery & Gosling, 2005).

CDSSs are computer-based systems that provide physicians with information or advice, including clinical knowledge, warnings and recommendations, in order to improve their clinical judgement and decision making. The researchers included a variety of systems ranging from electronic information resources like PubMed or Google, which loosely can be considered CDSSs - to specific diagnostic systems or electronic prescribing systems. Overall, consulting these systems was shown to improve the quality of clinical decisions and diagnoses (Garg et al., 2005). However, this general benefit can be offset at least to some extent by consequences of too much reliance on the CDSS when the suggestions or recommendations are wrong or suboptimal, reflecting automation bias. This negative effect has been shown by a number of studies comparing the correctness of clinical diagnoses before and after consulting a CDSS (Friedman et al. 1999; McKibbin & Fridsma 2006; Westbrook et al., 2005). Correct CDSS advices increased the number of correct diagnoses after consultation of the aid by 2-21%. However, at the same time, incorrect advice led clinicians in 6%-11% of cases to commit commission errors, that is, to change their initially correct diagnoses into wrong ones after consulting the false advice of the aid, which reduced the net gain effects of the CDSS significantly.

While these instances of automation bias involved only commission errors, more recent studies investigating the effects of electronic prescribing systems have revealed increased risks for both omission and commission errors caused by using this sort of CDSS (Lyell et al., 2017). The main function of electronic prescribing systems is to alert and warn a physician of possible risks due to an inappropriate or dangerous medication or unwanted interaction effects between different prescriptions. Compared to a control group that made prescription decisions without the support of the CDSS, correct alerts by the system successfully decreased the rate of prescription errors by about 43%. However, in case of missed alerts or false alarms of the CDSS, prescription errors due to omission errors (e.g. overlooking a wrong dose or a risky interaction effect which was not indicated by the system) and commission errors (e.g. abstaining from prescribing an actually safe medicine which was falsely indicated as risky by the system) increased by 28.7% and 56.9%, respectively. Participants who made omission errors in this study reported lower cognitive load compared with those who did not, suggesting that the allocation of insufficient

cognitive effort to the prescription task was a factor in these automation bias errors (Lyell, Magrabi, & Coiera, in press).

Another set of studies investigated possible risks of automation bias in interaction with computer-aided detection aids (CADs) in radiology (Alberdi, Povyakalo, Strigini & Ayton, 2004; Alberdi, Povyakalo, Strigini, Ayton & Given-Wilson, 2008). Experienced radiologists examined a set of mammograms either with or without the support of a detection aid that suggested areas containing lesions. The first study mainly focused on the risk of omission errors in cases where the CAD failed to prompt critical areas, that is, left a mammogram unmarked although there were signs of cancer or misplaced a prompt away from an area that was critical (Alberdi et al., 2004). Compared to a film reading without the aid, such false prompts decreased the detection rates considerably from 46% to 21% and 66% to 53%, respectively. Obviously, the film readers in this study tended to take the absence of a computer prompt as evidence for the absence of cancer. The authors interpreted this finding as evidence for automation bias, possibly due to a reduced vigilance when supported by a CAD. The second study looked specifically at performance consequences of false positive prompts, i.e. prompts that were placed although actually no signs of cancer were present (Alberdi et al., 2008). Again some evidence, for automation bias was found. Falsely placed prompts significantly raised the probability by 12.3% that the prompted areas actually were marked as malignancy compared to an unaided condition. Compared to the performance consequences of missed prompts, however, this effect was relatively weak, suggesting that with such systems false positive prompts providing opportunities for commission errors might be more tolerable than false negative prompts possibly provoking errors of omission.

Process Control. Some even stronger effects of automation bias have been reported from studies investigating performance consequences of automated decision-aids in simulated process control tasks (Bahner, Hueper & Manzey, 2008; Manzey, Reichenbach & Onnasch, 2012; Reichenbach, Onnasch & Manzey, 2011; Sauer, Chavaillaz & Wastell, 2016; Wickens, Gutzwiller, & Santamaria, 2015). In all of these studies a process-control microworld was used which simulates an autonomously running life-support system in a remote space habitat (AutoCAMS; Manzey et al., 2008). As a rule, students with a

background in engineering served as operators in these studies. Their task was to monitor the system and to intervene manually in case of failures. While performing this supervisory-control task they were supported by an Automated Fault Identification and Repair Agent (AFIRA) which, in case of failures in the life-support systems, provided the operators with an alert accompanied by an automatically generated failure diagnosis and/or further hints for proper failure management. In case of correct diagnoses and advice provided by AFIRA, operators' performance in terms of failure identification and fault management improved and became close to perfect, compared to when they had to perform the task without the decision aid. However, in case of an incorrect diagnosis provided by the aid, converging evidence for automation bias was found in all studies referred to above. This was reflected in up to 80% of operators committing a typical commission error, i.e. following the AFIRA advice although inspecting other available information would have proved it wrong. The effect was particularly evident for the first occurrence of such failure of the decision-aid ("first-failure effect"; Wickens, Clegg, Vieane, & Sebok, 2015). More specific analyses revealed that this sort of automation bias often occurred despite the fact that the operators had inspected other relevant data in order to cross-check the diagnosis of the aid at hand. That is, they either discounted this contradictory information or did not process it properly (Manzey et al., 2012).

Command and control. Finally, issues of automation bias have also been reported from interactions with intelligent decision support systems for command-and-control operations in the military domain (e.g., Crocoll & Coury, 1990; Cummings, 2004; Rovira et al., 2007). Decision-support systems in this domain include, among others, automated target identification and engagement systems. As reported by Cummings (2004), erroneously following such automated aids has already contributed to fatal military decisions, including friendly fires during the Iraq War. Systematic research in this domain has addressed the effects of front-end and back-end support on military decision-making and the risks if the decision-aids provided inaccurate advice (e.g., Crocoll & Coury, 1990; Rovira et al., 2007). For example, Rovira et al. (2007), investigated how different target-identification systems would impact the quality of military decisions under time pressure. The DSSs used in this study differed with respect to their reliability and to

what extent they just provided front-end information analysis support or also back-end decision support. In line with the results from other domains, all of the aids led to performance improvement in terms of reduced decision times and higher number of correct decisions made compared to an unaided control conditions. However, in case of incorrect advice of the aids a number of commission errors occurred and decision accuracy declined from 89% in the unaided condition to a mean of only 70% in the supported conditions. The most pronounced performance decrements were found for aids that provided high level of back-end support (i.e. made a specific recommendation for one particular response option) with an overall high level of reliability.

Factors Impacting Automation Bias.

It does little good to remind human operators that automation is not always reliable or trustworthy when their own experience tells them it can be trusted to perform correctly over long periods of time. (Billings, 1996, p. 97)

A number of factors have been suggested to impact the management of errors in interaction with automation and, thus, also the risk of automation bias when using DSSs. Major factors include individual characteristics and attitudes of the human operator, characteristics of system design, and different aspects of the task context (see also McBride, Rogers & Fisk, 2013 for an elaborated discussion of error management in human-automation interaction).

Individual factors

Trust and Overtrust. A key component in human-automation interaction is the extent to which the operator trusts the system (e.g., Chen & Barnes, 2014; Lee & See, 2004). Bailey and Scerbo (2007), for example, found an inverse relationship between trust and monitoring performance. In general, findings across many studies suggest that automation reliability is strongly associated with trust development and maintenance, and that operators adjust their trust in automation in line with its performance (Hoff & Bashir, 2015). Experience with reliable automated systems increases trust; negative experiences with automation can reduce it, with the negative feedback loop causing considerably stronger and longer lasting effects (Manzey et al., 2012). It seems reasonable to posit that the level of individual trust in a

certain DSS also impacts automation bias. This might also explain in part why professionals working with highly reliable systems are equally as susceptible to automation bias as novices (e.g., Mosier et al., 1998) - extensive experience with highly reliable automation induces trust and reduces the odds that experts will conduct a thorough information search to verify automated information and advice. However, clear empirical evidence that explicitly supports a link between individual trust and the occurrence of omission and commission errors is still lacking.

Furthermore there is evidence that individuals differ in their proneness to rely too much on automated systems. This individual tendency has been referred to as *complacency potential* (Singh, Molloy & Parasuraman, 1993) and is assumed to be dependent on perceived system properties (e.g. reliability), as well as individual characteristics of the human operator such as general attitudes towards technology or personality characteristics such as self-efficacy (Parasuraman & Manzey, 2010; Prinzel, 2002). More recent work even suggests that the basis of these individual differences is a genetic variance in the dopamine beta hydroxylase gene (Parasuraman, de Visser, Ming-Kuan & Greenwood (2012). However, more research is certainly needed before any decisive conclusions about systematic individual differences in proneness toward automation bias can be made.

Accountability for decisions. Early attempts to mitigate automation bias drew from the debiasing techniques of the social psychology literature. Mosier, Skitka and colleagues demonstrated that the imposition of pre-decisional accountability (before performing or making a decision; Hagafors & Brehmer, 1983) for accuracy and overall performance resulted in lower rates of omission and commission errors compared non-accountable student participants (Skitka, Mosier, & Burdick, 1999). Experimentally-manipulated accountability demands did not have the same impact for professional pilots in flight scenarios; however, pilots who reported an internalized perception of ‘accountability’ for their performance and strategies of interaction with automation were significantly more likely to double-check automated cues against other information and less likely to commit errors than those who did not share this perception (Mosier et al., 1998). The researchers suggested that ‘debiasing’ decision making through externally imposed accountability does not show much promise for eliminating automation bias in

professionals; it may be better to train behaviors that can mitigate its impact, such as verifying automated information against other available data. Other interventions, such as providing practical experiences with automation errors during training, may be more effective (Bahner et al., 2008).

System properties

Beside factors related to the human user of DSS, characteristics of the DSSs themselves, such as their reliability, whether they provide front-end or back-end support, their transparency and understandability, and how easy it is to cross-check or verify their outputs, have been suggested to impact the risk of automation bias.

Reliability. Automation bias is a direct reflection of what has been referred as ironies of automation (Bainbridge, 1983), that is, that highly reliable automation improves performance if it works correctly but may make things even worse compared to no automation support in case of automation errors. More specifically, the likelihood of automation bias seems to be directly linked to the reliability of DSSs. Direct supporting evidence for this relationship has been provided by Bailey and Scerbo (2007), using a complex system monitoring task, supported by different alarm systems as examples of simple front-end DSSs. They found that omission errors as consequence of automation errors (misses of detecting a critical state) increased from 32.4% to 48.3% when the reliability of the DSSs increased from 0.87 to 0.98. Similar results were also reported from Rovira et al. (2007). Using an automation monitoring task supported by different DSSs providing front-end or back-end support (see also below) they found overall higher levels of omission errors with reliable (.75) than unreliable (.50) DSSs. Furthermore, the predictability of reliability in terms of its dynamic variability seems to play a role with respect to omission errors. This is suggested by the study Parasuraman et al. (1993), which showed that omission errors in interaction with imperfect alarm systems were more likely when the reliability of the alarm system remained constant across experimental blocks than when it varied between low and high in different blocks. However, less and somewhat inconsistent data are available for direct links between DSS reliability and the occurrence of commission errors in case of wrong DSS recommendations. As reported above, Rovira et al. (2007) found higher rates of commission errors with a reliable (.80) than unreliable (.60) back-end DSS in a

command and control task, which corresponds to the findings for omission errors described above. Other studies also reported higher compliance and smaller rates of cross-checking the automation with higher reliability alarm systems but this seemed to reflect more of a well adapted behavior than a form of overcompliance and automation bias (Manzey et al., 2014). Studies that investigate compliance rates with complex DSSs or vary reliability and possible performance consequences in terms of automation bias are still lacking.

Front-end vs. back-end support. Some of the evidence for the possible significance of front-end versus back-end support in command and control tasks has been mentioned above. When studying risks of automation bias in interaction with automated target identification systems, Rovira et al. (2007) found that particularly systems providing back-end support reduced the overall reliability of decisions made in cases of wrong advice, which confirmed earlier findings of Crocoll and Coury (1990). Similar effects have also been reported from studies in aviation. For example, Sarter and Schroeder (2001) investigated the performance consequences of inflight-icing alerting systems in airplane cockpits. Two systems were compared in this study. The first one included a status display providing pilots (front end) information about the icing condition (wing icing or tailplane icing) but leaving the decision about a proper action with the pilots. The second system included a command display, directly providing (back-end) specific advice to the pilots what to do. The results showed that both sorts of DSS improved the overall speed and number of correct decisions made by the pilots considerably. However, in case of inaccurate advice, the systems caused more decision errors than in an unaided control condition and this effect was stronger for the back-end than the front-end aid. Laboratory studies addressing the impact of different types of DSSs in process control did not find differences between a front-end and back-end DSS in terms of manifest commission and omission errors. However, individuals spent less effort in cross-checking information against other available sources, that is, they showed monitoring behavior indicative of automation bias, when supported by a back-end than a front-end DSS (Manzey et al., 2012). Furthermore, there is evidence that the extent to which back-end support is provided might make a difference with respect to risks of automation bias in case of incorrect advice (Layton, Smith, & McCoy, 1994). In a study addressing the

possible performance consequences of a DSS supporting flight planning of pilots in aviation, the authors compared electronic flight planning tools which provided pilots with automatically generated suggestions for an optimum flight plan. They found that these tools do not always improve the quality of flight planning. Specifically, they found that pilots tended to accept the electronic flight plans even though they were suboptimal. Most important in this context, this tendency was stronger for tools recommending a specific flight plan (high level back-end support) than for tools that created different options but left the final choice to the pilots (low level back-end support).

Altogether, the different studies referred to above suggest that front-end DSSs might be less conducive to automation bias in case of incorrect advice than more directive back-end DSSs, and that the more definitive the advice provided by the latter the more conducive to automation bias they are. This fits to the more general “lumberjack” hypothesis stated by Onnasch et al. (2014) that higher degrees of automation provide higher benefits than lower levels in cases of reliable routine operation but increases risks of errors and return-to-manual performance decrements in cases of automation failures. However, the effects with respect to automation bias found in the different studies were usually relatively small and other design factors of DSSs might be even more important for affecting risks of automation bias.

Automation visibility. Such other design factors might include the transparency of DSSs, the understandability of their functioning, and the effort it takes to cross-check and verify their outputs. Specifically, system understandability and transparency have been considered to be key antecedents to the development and calibration of system trust. (e.g., Chen, Barnes, & Harper-Sciarini, 2011; Lyons et al, 2016; Sheridan, 1988). A recent study in the railway sector, for example, identified understanding of the automation, rather than system reliability, as the key component in developing trust in a DSS (Balfe, Sharples, & Wilson, 2018). Evidence for the importance of visibility in mitigating automation bias has been provided by McGuirl and Sarter (2006). Capitalizing on the earlier research on effects of inflight-icing alerts (Sarter & Schroeder, 2001) they investigated how pilots’ decision making in response to inflight icing encounters would be affected by DSSs that provided system confidence information together with the automatically generated front-end or back-end information/advice. The confidence

information was updated on a trial-by-trial basis and was presented to the participants in a separate trend display. Independent of whether front-end or back-end support was provided, the pilots receiving this additional information were less prone to automation bias in case of inaccurate information, obviously because they were better able to assess the validity of the aids' recommendations. Analogous effects have recently been reported from so called likelihood alarm systems (front-end DSS), which provide better options to estimate the likelihood of critical events in case of alerts than binary alarm systems and which were found to improve human decision making particularly in response to system errors, i.e. false or missed alerts (Wiczorek & Manzey, 2014). However, independent of whether or not likelihood information was provided, the fewest number of commission and omission errors were made when information to cross-check and verify the output of the alarm systems was available and easily accessible. This suggests that also the accessibility of automation verification information can impact the risk of automation bias. Thus, it can be expected that design decisions concerning the display of verification data (e.g. surface display vs. layered or hidden display) will impact whether or not decision makers will trust DSSs or make the effort to verify their output. Overall, automation visibility, that is, its transparency in terms of its sources and functioning (Dorneich et al., 2015) and the ease of accessibility of automation verification information, seems to be a key determinant in whether decision makers will cross-check or just (over-)trust system information. This directly fits to recent findings from a meta-analysis of automation bias in interaction with DSSs in different domains, which also suggests that risks of automation bias are directly related to the effort needed to verify the output of a DSS (Lyell & Coiera, 2017).

Task Context

DSSs are introduced to support individuals in accomplishing certain tasks. Thus, it seems reasonable to assume that the task itself or the specific situational circumstances under which a task has to be performed, that is the task context, also impacts how individuals interact with DSSs and make use of them. For example, it can be assumed that factors like time-pressure, workload, the expected performance consequences of decision errors, or the social context in which a task has to be performed might have an

impact on the risk of automation bias. However, only few studies have addressed these sorts of factors, thus far.

Time pressure. It has been known for some time that time pressure can have a considerable impact on human judgment and decision making (Edland & Svenson, 1993). More specifically, humans have been shown to compensate for time-pressure in judgment and decision making by abandoning a time consuming rational analytic strategy, based on comprehensive information search, in favor of a faster heuristic strategy based on considering only most important (valid) cues (Rieskamp & Hoffrage, 2008). Often, this use of heuristics instead of full information search and weighting allows for making quicker decisions without compromising quality to a significant extent (Gigerenzer & Gaissmaier, 2011).

Analogous to these findings, one might assume that time-pressure also would lead humans to depend more on DSSs, that is, to use their particularly salient and usually (but not always) correct advice directly as basis for fast judgments and decision-making without fully evaluating the available information from the environment. This assumption has been addressed in a series of experiments by Rice and colleagues (Rice, Hughes, McCarley & Keller, 2008; Rice & Keller, 2009; Tunstall, Rice, Mehta, Dunbar & Oyman, 2014). Specifically, they investigated to what extent time-pressure would affect dependence on front-end DSSs in situations where humans had to make binary decisions about the presence of critical targets based on the visual inspection of complex images. After getting the advice of the DSS (i.e., critical target present or absent), participants in their studies had either two or eight seconds for own inspection of the visual images before they had to commit to a decision. The DSSs used in these studies did not miss any targets but provided a different number of false alarms affording the opportunity for automation bias in terms of commission errors. What Rice et al. consistently found was a higher compliance rate with the DSS under the high compared to low time-pressure conditions. This effect emerged largely independent of the reliability of the DSSs ranging from highly reliable (.95) to fairly reliable (.65) systems. At least part of this higher compliance with DSS advice under time-pressure seems to reflect fairly rational behavior, given that the time for comprehensive information search is limited and the objective dependence on an DSS gets higher. This was also reflected in the performance effects. The higher

compliance rates under time-pressure did not lead to overall performance decrements when the reliability of the DSS was high. Compared to an unsupported control condition, even more correct decisions were produced under time-pressure with support of the highly reliable DSS.

It seems that in this case the performance costs of automation bias effects in terms of commission errors when erroneously responding to false alarms were less than the performance benefits gained by also following the DSS more often when it was correct. As a consequence, Rice et al. recommend moderate time-pressure as a possible means to improve compliance with highly reliable DSSs. However, clear overall performance costs due to automation bias effects emerged with the least reliable DSS. Here the same strategy of higher dependence on the DSS under time-pressure caused performance to drop even below the performance achieved when no DSS was available (Rice et al., 2008; Rice & Keller, 2009). These results directly indicate the mixed blessing of heuristic use of a DSS under conditions of time-pressure. It can improve performance if the reliability of the DSS is high but can lead to severe automation bias errors in case of comparatively low DSS reliability. Surprisingly, the empirical evidence for the impact of time pressure is largely limited to front-end DSSs. A generalizability to back-end DSSs seems highly plausible, but requires further evidence.

Workload. Similarly, also the overall workload of an individual has repeatedly shown to affect an individual's reliance and compliance in interactions with DSSs. An example of the former is an early finding of Parasuraman et al. (1993) suggesting that complacency effects and resulting omission errors in response to unreliable automation particularly arise in multitasking contexts, i.e. when the task supported by an DSS is one of several tasks to be performed concurrently. It seems that when multitasking, individuals tend to lower the priority of the task supported by a DSS and the effort to be invested in cross-checking the automation in favor of the other tasks. Whereas such behavior, on the one hand, certainly reflects a reasonable (and intended) consequence of introducing a DSS (Moray, 2003) it also can increase the risk of automation bias if it is not balanced properly.

Other findings have provided converging evidence of the same processes occurring for errors of commission. For example, Manzey et al. (2014; Exp. 2 and 4) compared compliance rates with front-end

DSSs (alarms systems) under conditions where participants had to perform one or two tasks concurrently to the task supported by the DSS. The raised workload in the two-task condition increased the direct compliance rate, that is the frequency of responses to the alarm without consulting other available information, for a relatively unreliable (.70) DSS from about 20% to almost 60%, resulting in a considerably higher rate of commission errors when the workload was high. Even more direct support for workload impacting the dependence on DSSs has been provided by other studies (Biros, Daly & Gunsch, 2004; Dorneich et al., 2015). Biros et al. (2004) studied to what extent workload would change the link between trust in DSSs and reliance on their advice in a command and control scenario. More specifically, they varied trust in a DSS supporting tactical decisions by means of different instructions and investigated to what extent this variation would have an impact on DSS use depending on the overall task load. Task load was operationally defined by the complexity of information to be considered in the decision-task. While differences in trust directly predicted the use of automation in the low task-load condition, participants tended to rely on the automation independent of their trust in the DSS when the task-load was high. The higher task-load obviously led to less skeptical over-reliance in the DSS although the participants were aware that the advice might be wrong and, thus, accepted the higher risk of committing commission errors when using the aid. Similarly, Dorneich et al. (2015) found that air transport pilots tended to over-trust an information automation system when they were under high workload and chose the top plan suggested by the system, even though information was missing and the plan was not the best one. Note that these findings parallel the findings of time-pressure effects on human dependence on DSSs. Particularly multitasking situations, but also high workload situations in general can be considered as situations where the time for judgment and decision-making is lowered due to other task demands, implicitly inducing a sort of time-pressure for humans.

However, also workload that is too low (“underload”) might also be detrimental for human-automation performance and can increase automation bias effects. This is suggested by the recent work of Lyell and colleagues, who found that participants who made omission errors reported lower cognitive load compared with those who did not, suggesting that low cognitive load led to the allocation of

insufficient cognitive effort to the experimental task, which was a factor in these automation bias errors (Lyell et al., in press). Yet, it is not clear what cause and effect is in these findings and clearly more research on effects of underload will be needed before any decisive conclusions about effects of underload on automation bias can be drawn.

Consequences of Errors. Evidence that the consequences of potential errors committed by an operator might affect verification behavior and the risk of automation bias was provided by two of the first studies of automation bias (Mosier et al., 1998; Mosier, Skitka, Dunbar, & McDonnell, 2001). In this research, pilots were required to perform different flight tasks during a simulated flight from San Francisco to Los Angeles (Mosier et al., 1998) or different approaches to San Francisco (Mosier et al., 2001) in a low-fidelity part-task flight simulator. During the flight different automation failures occurred which provided opportunities to commit omission and commission errors as manifestations of overreliance or overcompliance with the automated systems. In both studies, the criticality of errors turned out to be an important predictor of omission errors. For example, automation failures related to altitude control (e.g. mis-loading an altitude clearance in the autopilot) were more often detected than failures related to the communication system (e.g. mis-loading a frequency change). Thus, it seems that pilots still monitored and verified actions of DSS related to flight critical tasks but were less vigilant in monitoring subsystems less critical for flight safety, and this effect occurred independent of whether the simulations were accomplished by single pilots (Mosier et al., 1998) or two-person crews (Mosier et al., 2001).

Individuals vs. teams. Most of the research on the automation bias phenomenon has been conducted in a single-person performance configuration. However, some highly automated decision environments, such as aircraft cockpits, involve teams rather than solo decision makers. Whether or not this condition of shared responsibility makes a difference with respect to automation bias and risks of committing omission and commission errors when DSSs err was investigated in a pair of studies by Mosier and colleagues. Specifically, they examined individuals vs. teams of commercial class cockpit pilots (Mosier et al., 2001) and students (Skitka, Mosier, Burdick, & Rosenblatt, 2000b) performing simulated flight tasks under

varying instruction conditions. Results demonstrated the persistence of automation bias in teams for both samples.

Conclusion, Present Trends and Future Possibilities

Clearly, automation bias is still a ubiquitous phenomenon in the interaction of humans and DSSs which is particularly relevant with highly, but not perfectly reliable systems. It thus directly reflects an irony of automation (Bainbridge, 1983), namely that highly reliable DSSs on the one hand certainly improve the human judgement and decision making when working properly, but at the same time, lead to automation bias in case of automation errors.

Even more important, automation bias does not necessarily result from human weaknesses in automation monitoring but can directly result from quite rational behavior in interaction with automated systems (Moray, 2003). Thus, directly corresponding to the assumption that using heuristics in human decision making is not necessarily bad but can “make us smart” (Gigerenzer & Todd, 1999), using highly reliable DSSs as a heuristic most times leads to positive effects like reduced cognitive effort and better decisions. And most organizations *want* their decision makers to trust – and not constantly second-guess - the automated support they have provided! Luedtke and Moebus (2005) have proposed that interactions with highly reliable DSS over time might lead to an effect of *learned carelessness*. That is, making the repeated experience that the automation works properly even without close monitoring, the tendency to directly use the advices of DSSs as a heuristic for decision-making is amplified by a sort of positive feedback loop (see also Parasuraman & Manzey, 2010). This might also explain why human-focused approaches to improve automation verification and mitigating automation bias, such as making individuals accountable for their interactions with automation or aware of the non-perfect reliability of systems and the related risk of automation bias, have had minimal success (Skitka, Mosier & Burdick, 2000a), unless human users have the direct experience of automation errors (Bahner et al., 2008). Thus, the main challenge of mitigating automation bias lays with a proper design of DSSs, specifically in finding a delicate balance between designing automation that facilitates decision making and creating an

environment that fosters automation bias. Achieving this balance is essential to creating a human-automation match that improves performance considerably without introducing new risks, even though it still may not be as perfect as a ‘made in heaven.’ A few current design trends and future possibilities are geared toward accomplishing this.

Balancing the Tradeoffs

Several design factors which might help to mitigate the risk of automation bias in interaction with DSS can be identified from the research presented above. Beside the tendency for front-end DSSs to be a bit more resilient toward automation bias than more directive back-end DSSs, a more general take-away message from the research discussed here concerns the role of transparency including understandability and predictability. Particularly, proper system feedback during use – especially cognitive feedback that provides information on how a DSS functions and the relationships among information and DSS output – is a form of transparency that can facilitate understanding of DSS functioning and accurate calibration for its use (Seong & Bisantz, 2008). Many early DSSs were opaque in their functioning and did not provide the decision maker with any map of their situation (mental) model, rationale for their recommendations, or justification for their recommendations. Early in the evolution of aviation automation, Billings (1996) already emphasized the need to train pilots *how systems operate* rather than simply *how to operate systems*: “If a pilot does not have an adequate internal model about how the computer works when it is functioning properly, it will be far more difficult for him or her to detect a subtle failure. We cannot always predict failure modes in these more complex digital systems, so we must provide pilots with adequate understanding of how and why aircraft automation functions as it does” (p. 96). Operators must have sufficient knowledge of what DSSs can do, what they “know,” and how they function within the context of other systems as well as knowledge of their limitations, in order to utilize them efficiently and avoid errors of automation bias. System feedback that seems to be particularly important in this respect is the provision of confidence information along with advice and recommendations provided by DSS. Examples include information about the quality of the data base used by a DSS (e.g. the accuracy of positions identified by a GPS), likelihood information about the presence of a critical situation indicated

by a front-end DSS (Wiczorek & Manzey, 2014), or feedback about the strength of evidence supporting the output of a back-end DSS (McGuirl & Sarter, 2006).

DSS as a Team Member

An overarching design principle is to look at DSSs as members of a human-automation team. The team-member metaphor as a possible guiding principle has been introduced to automation research and design by Christoffersen and Woods (2002). With respect to DSSs this suggests to design a system to function as like a true team member, under the assumption that components of successful human-automation teams mirror those of effective human-human teams (Mosier, Fischer, Burian, & Kochan, 2017). This perspective is consistent with current tendencies to adapt human-human concepts – for example, trust – to human-automation interaction and to evaluate automated DSSs according to these human characteristics. The ideal DSS then would incorporate not only the most needed abilities of automated decision support – such as the ability to sense, synthesize, and integrate large amounts of data and information, to perform calculations quickly, and to detect malfunctions and failures – but also incorporate into the design characteristics that are desired in a human crew member and which are known to make a difference in human teamwork. On the one hand this would include the sort of transparency and understandability mentioned above, as ideal human crewmembers are observable when acting, are predictable with respect to their next actions and provide a justification of their intentions and actions. Beyond that, ideal DSS would also resemble human crew members in other aspects, including responsiveness to direction, flexibility in providing advices and recommendations on a level (front-end versus back-end) needed by the human, and the ability to adapt to contextual factors like workload and time-pressure which current DSSs may not be able to consider. Finally, ideal DSS would be able to monitor the human as the human monitors the DSS.

Concrete examples of more advanced DSSs in this respect include DSSs that are adaptable or even adaptive to the needs of human users, ensuring that the best type of support (front-end and/or back-end; low level vs. high level) is easily available and verifiable at the right time, especially when decisions need to be made under time pressure (Mosier et al., 2017; Burian, Mosier, Fischer, & Kochan, in press).

Adaptable DSSs are adjusted by the operator, who maintains control over automation and is able to designate whether the human or automation will do all or part of tasks. Thus, individual operators can choose the level of support they prefer. Examples how this might be implemented and used by individual operators are provided by Sauer and Chavaillaz (2018). In contrast, with *adaptive* systems, the automation controls the division of tasks, and may automatically change the sort of support provided in response to individual states of the human user (e.g. fatigue, stress) or task-context factors (e.g. time-pressure, workload; Sheridan & Parasuraman, 2006). An example of the latter is the TCAS (Traffic Collision Avoidance System) in modern airplanes which dynamically changes its basic characteristic from front-end to back-end support with increasing time-pressure. As long as the situation allows, TCAS provides front-end informational support with accompanying lower risks of automation bias, and changes to back-end action directives only when time-pressure becomes so high (i.e., imminent collision) that immediate compliance with the highly reliable conflict resolution advisory - despite associated risks of automation bias - is the safer option. Another conceptualization of adaptive automation is context-sensitive information automation, which would sense and take into account the situation or context, tailoring information support to characteristics such as specific task demands, environmental factors, system status, and human cognitive and performance variables (Mosier et al., 2017). This notion, however, is only at the conceptual stage, and the potential automation bias risks in this type of system are unknown.

A recent and even more far-reaching approach, enabled by current technological advances, indeed treats human and ‘automated agents’ as interdependent team members who share a common mental model of situations and goals and can coordinate and collaborate activities, monitor each other, provide feedback to each other, and adapt dynamically to contextual demands. This approach, characterized as Coactive Design (e.g., Bradshaw, Dignum, Jonker, & Sierhuis, 2012; Johnson, et al., 2011), adheres to the design principles of observability, predictability and directability (Johnson et al., 2014), which is intended to facilitate teamwork behaviors such as monitoring progress and providing back-up behavior, and make coordinated action possible. The implementation of Coactive Design is still in its infancy,

although it has already been applied to the development of a humanoid robot to assist a human operator during disaster relief (Johnson, Bradshaw, Feltovich et al., 2014). However, the concept of human-DSS interdependence and automation autonomy will most likely figure in the design of future automated decision support. One caveat for the notion of a more human-like automated team member - because ‘humanizing’ automation (particularly anthropomorphism) enhances trust, designers will have to ensure that it does not inadvertently promote automation bias (e.g., de Visser et al., 2016; Pak, Fink, Price, Bass, & Sturre, 2012)

In summary, it seems that the match between human and DSSs has many advantages but is not perfect. Actually, it seems that humans and DSSs represent more a match made on earth than a match made in heaven! As with all other earth-made matches, continuous attention is needed from both sides, the system design and the human user, to keep the match working smoothly, and to achieve the best balance between maximizing the benefits of technological advances and, at the same time, minimizing the risks of automation bias. The present review suggests that this is a complex undertaking which needs consideration of system characteristics, characteristics of the human user and characteristics of the situational context in which DSSs are used.

References

- Alberdi, E., Povyakalo, A., Strigini, L., & Ayton, P. (2004). Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography. *Academic radiology*, 11(8), 909-918.
- Alberdi, E., Poviakalo, A.A., Strigini, L., Ayton, P., & Given-Wilson, R. (2008). CAD in mammography: lesion-level versus case-level analysis of the effects of prompts on human decisions. *International Journal of Computer-Assisted Radiology and Surgery*, 3, 115-122.
- Bahner, E., Hueper, A.-D., & Manzey, D. (2008). Misuse of automated decision aids: Complacency, automation bias and the impact of training experience. *International Journal of Human-Computer Studies*, 66, 688-699.
- Bailey, N. R., & Scerbo, M. W. (2007). Automation-induced complacency for monitoring highly reliable systems: The role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science*, 8(4), 321-348.
- Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775-779.
- Balfe, N., Sharples, S., & Wilson, J. R. (2018). Understanding is key: An analysis of factors pertaining to trust in a real-world automation system. *Human Factors*, 60, 477-495.
- Billings, C. E. (1996). *Human-Centered Aviation Automation: Principles and Guidelines*. NASA TM 110381. Moffett Field, CA: NASA Ames Research Center.

- Biros, D.P., Daly, M., & Gunsch, G. (2004). The influence of task load and automation trust on deception detection. *Group Decision and Negotiation*, 13, 173-189.
- Bradshaw, J. M., Dignum, V., Jonker, C., & Sierhuis, M. (2012). Human-agent-robot teamwork. *IEEE Intelligent Systems*, 27(2), 8-13.
- Burian, B. K., Mosier, K. L., Fischer, U. M., & Kochan, J. A. (in press). New teams on the flight deck: Humans and context-sensitive information automation. In E. Salas & F. Jentsch (Eds.) *Human Factors in Aviation* (3rd edition). New York: Academic Press.
- Chen, J. Y., & Barnes, M. J. (2014). Human-agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems*, 44 (1), 13-29.
- Chen, J. Y., Barnes, M. J., & Harper-Sciarni, M. (2011). Supervisory control of multiple robots: Human-performance issues and user-interface design. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 41(4), 435-454.
- Christoff ersen, K. & Woods, D. D. (2002).How to make automated systems team players. In E. Salas (Ed.), *Advances in human performance and cognitive engineering research. Vol. 2. Automation* (pp. 1–12). Burlington: Elsevier.
- Crocoll, W.M., & Coury, B.G. (1990). Status or recommendation: selecting the type of information for decision aiding. *Proceedings of the Human Factors Society 34th Annual Meeting* (pp. 1525-1528). Santa Monica: Human Factors and Ergonomics Society.
- Cummings, M. (2004, September). Automation bias in intelligent time critical decision support systems. In *AIAA 1st Intelligent Systems Technical Conference* (p. 6313).
- Cummings, M.L. (2004). *Automation bias in intelligent time critical decision support systems*. Paper presented to the AIAA 1st Intelligent Systems Technical Conference, September 2004. [available from: [web.mit.edu/aeroastro/labs/halab/papers/ CummingsAIAAbias.pdf](http://web.mit.edu/aeroastro/labs/halab/papers/CummingsAIAAbias.pdf); 28 Feb 2010]
- de Boer, R. J., Heems, W., & Hurts, K. (2014). The duration of automation bias in a realistic setting. *International Journal of Aviation Psychology*, 24, 287-299.
- de Visser E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, 22(3), 331-349.
- Dorneich, M. C., Dudley, R., Rogers, W., Letsu-Dake, E., Whitlow, S. F., Dillard, M., & Nelson, E. (2015). Evaluation of information quality and automation visibility in information on the flight deck. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 59, 284-288. Santa Monica, CA: HFES.
- Dutch Safety Board. (2010). *Crashed during approach, Boeing 737-800, near Amsterdam Schiphol airport, 25 February 2009*. The Hague, the Netherlands.
- Edland, A. & Svenson, O. (1993). Judgement and decision making under time pressure. Studies and findings. In O. Svenson & A. Maule (eds,) *Time pressure and stress in human judgement and decision making* (pp. 27-40). New York: Plenum Press.
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 85–104.
- Endsley, M. R. (2000). Theoretical underpinnings of situational awareness: A critical review. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness, analysis, and measurement* (pp. 317-341). Mahwah, NJ: Erlbaum.
- Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42, 462–492.
- Friedman, C.P., Elstein, A.S., Wolf, F.M., Murphy, G. C., Franz, T.M., Heggerling, P.S., Fine, P.L., Miller, T.M., & Abraham, V. (1999). Enhancement of clinicians' diagnostic reasoning by computer-based consultation: a multiside study of 2 systems. *Journal of the American Medical Association*, 282, 1851-1856.

- Garg, A.X., Adhikari, N.K., McDonald, H., Rosas-Arellano, P.M., Devereaux, P. J., Beyene, J., Sam, J., & Haynes, R.B. (2005). Effects of clinical decision support systems on practitioner performance and patient outcome: a systematic review. *Journal of the American Medical Association*, 293, 1223-1238.
- Gigerenzer, G. & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62, 451–82
- Gigerenzer, G., & Todd, P.M. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Goddard, K., Roudsari, A., & Wyatt, J.C. (2012). Automation bias: a systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, 19, 121-127.
- Goddard, K., Roudsari, A., & Wyatt, J.C. (2014). Automation bias: Empirical results assessing influencing factors. *International Journal of Medical Informatics*, 83, 368-375.
- Hagafors, R., & Brehmer, B. (1983). Does having to justify one's decisions change the nature of the decision process? *Organizational Behavior and Human Performance*, 31, 223-232.
- Hoff, K. A., & Bashir, M. (2015). Trust in automation integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3), 407-434.
- Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., van Riemsdijk, M. B., & Sierhuis, M. (2011). The fundamental principle of coactive design: Interdependence must shape autonomy. In M. De Vos, N. Fornara, J. Pitt, & G. Vouros (Eds.), *Coordination, organizations, institutions, and norms in agent systems VI* (Vol. 6541, pp. 172–191). Springer Berlin Heidelberg.
- Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., van Riemsdijk, M. B., & Sierhuis, M. (2014). Coactive design: Designing support for interdependence in joint activity. *Journal of Human-Robot Interaction*, 3(1), 43-69.
- Kahnemann, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus & Giroux.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. NY: Cambridge University Press.
- Layton, C., Smith, P. J., & McCoy, C. E. (1994). Design of a cooperative problem-solving system for en-route flight planning: An empirical evaluation. *Human Factors*, 36, 94-119.
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50-80.
- Luedtke, A., & Moebus, C. (2005). A case study for using a cognitive model of learned carelessness in cognitive engineering. In G. Salvendy (Ed.), *Proceedings of the 11th International Conference of Human-Computer Interaction*. Mahwah, NJ: Erlbaum. Retrieved from https://www.uni-oldenburg.de/fileadmin/user_upload/informatik/ag/lks/download/Publikationen/2005/Luedtke_Moebus_crv.pdf <26 June, 2018>
- Lyell, D., & Coiera, E. (2017). Automation bias and verification complexity: a systematic review. *Journal of the American Medical Informatics Association*, 24(2), 423-431.
- Lyell, D., Magrabi, F., & Coiera, E. (in press). The effect of cognitive load and task complexity on automation bias in electronic prescribing. *Human Factors*
- Lyell, D., Magrabi, F., Raban, M.Z., Pont, L.G., Baysari, M.T., Day, R.O., & Coiera, E. (2017). Automation bias in electronic prescribing. *BMC Medical Informatics and Decision Making*, 17, 28. <https://doi.org/10.1186/s12911-017-0425-5>.
- Lyons, J. B., Koltai, K. S., Ho, N. T., Johnson, W. B., Smith, D. E., & Shively, R. J. (2016). Engineering trust in complex automated systems. *Ergonomics in Design*, 24(1), 13-17.
- Manzey, D., Bleil, M., Bahner-Heyne, J.E., Klostermann, A., Onnasch, L., Reichenbach, J., & Röttger, S. (2008). *Auto-CAMS 2.0 manual*. Retrieved from <http://www.aio.tu-berlin.de/?id=30492>

- Manzey, D., Gerard, N. , & Wiczorek, R. (2014). Decision-making and response strategies in interaction with alarms: the impact of alarm reliability, availability of alarm validity information and workload. *Ergonomics*, 57, 1833-1855
- Manzey, D., Reichenbach, J., & Onnasch, L. (2012). Human performance consequences of automated decision aids: The impact of degree of automation and system experience. *Journal of Cognitive Engineering and Decision Making*, 6, 57 – 87.
- McBride, S.E., Rogers, W.A. & Fisk, A.D. (2013). Understanding human management of automation errors. *Theoretical Issues in Ergonomics Science*, 15, 545-577.
- McGuirl, J. M., & Sarter, N. B. (2006). Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information. *Human factors*, 48(4), 656-665.
- McKibbin, K.A., & Fridsma, D.B. (2006). Effectiveness of clinical-selected electronic information resources for answering primary care physician's information needs. *Journal of the American Medical Informatics Association*, 13, 653-659.
- Metzger, U., & Parasuraman, R. (2005). Automation in future air traffic management: Effects of decision aid reliability on controller performance and mental workload. *Human Factors*, 47, 35-49.
- Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors*, 46, 196–204.
- Moray, N. (2003). Monitoring, complacency, scepticism and eutactic behaviour. *International journal of industrial ergonomics*, 31(3), 175-178.
- Mosier, K. L., & Fischer, U. M. (2010). Judgment and decision making by individuals and teams: Issues, models and applications. In D. Harris (Ed.), *Reviews of Human Factors, Volume 6* (pp. 198-256). Santa Monica, CA: Human Factors and Ergonomics Society.
- Mosier, K., Fischer, U., Burian, B., & Kochan, J. (2017). *Autonomous, Context-Sensitive, Task Management Systems and Decision Support Tools I: Contextual Constraints and Information Sources*. NASA/TM-2017-219565. NASA Ames Research Center.
- Mosier, K.L., Palmer, E.A., & Degani, A. (1992). Electronic checklists: Implications for decision making. *Proceedings of the Human Factors Society 36th Annual Meeting* (pp. 7-11). Santa Monica: Human Factors Society.
- Mosier, K.L., & Skitka, L.J. (1996). *Human decision makers and automated decision aids: made for each other?* In: R. Parasuraman, & M. Mouloua (Eds.), *Automation and human performance: Theory and application*. (pp. 201–220). Mahwah, NJ: Erlbaum.
- Mosier, K.L., Skitka, L.J., Dunbar, M., & McDonnell, L. (2001). Aircrews and automation bias: The advantages of teamwork? *The International Journal of Aviation Psychology*, 11, 1-14.
- Mosier, K.L., Skitka, L.J., Heers, S., & Burdick, M. (1998). Automation bias: Decision-making and performance in high-tech cockpits. *The International Journal of Aviation Psychology*, 8, 47-63.
- Onnasch, L., Wickens, C.D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors*, 56, 476–488.
- Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics*, 55(9), 1059-1072.
- Parasuraman, R., & Manzey, D. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52, 381–410.
- Parasuraman, R. M., Molloy, R., & Singh I. L. (1993). Performance consequences of automation induced" complacency. *International Journal of Aviation Psychology*, 3, 1-23.
- Parasuraman, R., de Visser, E., Lin, M. K., & Greenwood, P. M. (2012). Dopamine beta hydroxylase genotype identifies individuals less susceptible to bias in computer-assisted decision making. *PloS one*, 7(6), e39675.

- Parasuraman, R., Sheridan, T., & Wickens, C. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-30, 286-297.
- Prinzel, L.J. (2002). *The relationship of self-efficacy and complacency in pilot-automation interaction*. NASA/TM-2002-211925. Hampton, Virginia: NASA Langley Research Center.
- Reichenbach, J., Onnasch, L., & Manzey, D. (2011). Human performance consequences of automated decision aids in states of sleep loss. *Human Factors*, 53, 717-728
- Rice, S., Hughes, J., McCarley J.S. & Keller, D. (2008). Automation dependency and performance gain under time pressure. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting* (pp. 1326-1329). Santa Monica: Human Factors and Ergonomics Society.
- Rice, S. & Keller, D. (2009). Automation reliance under time pressure. *Cognitive Technology*, 14, 36-44.
- Rieskamp, J. & Hoffrage, Inferences under time-pressure: How opportunity costs affect strategy selection. *Acta Psychologica*, 127, 258-276 .
- Rovira, E., & Parasuraman, R. (2010). Transitioning to future air traffic management: Effects of imperfect automation on controller attention and performance. *Human Factors*, 52, 411-425.
- Rovira, E., McGarry, K., & Parasuraman, R. (2007). Effects of imperfect automation on decision making in a simulated command and control task. *Human Factors*, 49, 76-87.
- Sarter, N.B., & Schroeder, B. (2001). Supporting decision making and action selection under time pressure and uncertainty: The case of in-flight icing. *Human Factors*, 43, 573-583.
- Sauer, J., & Chavallaiz, A. (2018). How operators make use of wide-choice adaptable automation: observations from a series of experimental studies. *Theoretical Issues in Ergonomics Science*, 19, 135-155.
- Sauer, J., Chavaillaz, A., & Wastell, D. (2016). Experience of automation failures in training: effects on trust, automation bias, complacency and performance. *Ergonomics*, 59, 767-780
- Seong, Y., & Bisantz, A. M. (2008). The impact of cognitive feedback on judgment performance and trust with decision aids. *International Journal of Industrial Ergonomics*, 38, 608-625.
- Sheridan, T. (1988). Trustworthiness of command and control systems. In *Proceedings of the International Federation of Automatic Control Symposium on Man-Machine Systems* (pp. 427-431). Elmsford, NY: Pergamon.
- Sheridan, T., & Parasuraman, R. (2006). Human-automation interaction. In R. S. Nickerson (Ed.), *Reviews of human factors and ergonomics* (Vol. 1, pp. 89-129). Santa Monica, CA: Human Factors and Ergonomics Society.
- Simons, D.J., & Chabris, C.F. (1999). Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception*, 28, 1059-1074.
- Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation-induced “complacency”: Development of a complacency-potential scale. *International Journal of Aviation Psychology*, 3, 111-122.
- Skitka, L.J., Mosier, K.L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies*, 51, 991-1006.
- Skitka, L.J., Mosier, K.L., & Burdick, M. (2000a). Accountability and automation bias. *International Journal of Human-Computer Studies*, 52, 701-717.
- Skitka, L.J., Mosier, K.L., Burdick, M., & Rosenblatt, B. (2000b). Automation bias and errors: Are crews better than individuals? *The International Journal of Aviation Psychology*, 10, 85-97.
- Tsai, T.L., Fridsma, D.B., & Gatti, G. (2003). Computer decision support as a source of interpretation error: the case of electrocardiograms. *Journal of the American Medical Informatics Association*, 10, 478-483.

- Tunstall, C., Rice, S., Dunbar, V. & Oyman, K. (2014). Time pressure has limited benefits for human-automation performance. *Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting* (pp. 1043-1046). Santa Monica: Human Factors and Ergonomics Society.
- Westbrook, J.I., Coirea, W.E., & Gosling, A.S. (2005). Do online information retrieval systems help experienced clinicians answer clinical questions? *Journal of the American Medical Informatics Association*, 12, 315-321.
- Wickens, C.D., Clegg, B., Vieane, A. Z., and Sebok, A.L. (2015). Complacency and automation bias in the use of imperfect automation. *Human Factors*, 57, 728-739.
- Wickens, C. D., Gutzwiller, R. S., & Santamaria, A. (2015). Discrete task switching in overload: A meta-analysis and a model. *International Journal of Human-Computer Studies*, 79, 79-84.
- Wiczorek, R. & Manzey, D. (2014). Supporting attention allocation in multitask environments: Effects of likelihood alarm systems on trust, behavior and performance. *Human Factors*, 56, 1209–1221
- Wiener, E. L. (1985). Beyond the sterile cockpit. *Human Factors*, 27, 75-90.